Thematic issue: Biologics in autoimmune diseases

Analyst, 137, 10, 2012, DOI: 10.1039/C2AN35122F

V. Pirro, L.S. Eberlin, P. Oliveri, R.G. Cooks
Volume 137, Royal Society of Chemistry, 2012, 2374-2380

Abstract

Desorption electrospray ionization (DESI) is an ambient mass spectrometry (MS) technique that can be operated in an imaging mode. It is known to provide valuable information on disease state and grade based on lipid profiles in tissue sections. Comprehensive exploration of the spatial and chemical information contained in 2D MS images requires further development of methods for data treatment and interpretation in conjunction with multivariate analysis. In this study, we employ an interactive approach based on principal component analysis (PCA) to interpret the chemical and spatial information obtained from MSimaging of human bladder, kidney, germ cell and prostate cancer and adjacent normal tissues. This multivariate strategy facilitated distinction between tumor and normal tissue by correlating the lipid information with pathological evaluation of the same samples. Some common lipid ions, such as those of m/z 885.5 and m/z 788.5, nominally PI(18 : 0/20 : 4) and PS(18 : 0/18 : 1), as well as ions of free fatty acids and their dimers, appeared to be highly characterizing for different types of human cancers, while other ions, such as those of m/z 465.5 (cholesterol sulfate) for prostate cancer tissue and m/z 795.5 (seminolipid 16 : 0/16 : 0) for germ tissue, appeared to be extremely selective for the type of tissue analyzed. These data confirm that lipid profiles can reflect not only the disease/health state of tissue but also are characteristic of tissue type. The manual interactive strategy presented here is particularly useful to visualize the information contained in hyperspectral MS images by automatically connecting regions of PCA score space to pixels of the 2D physical object. The procedures developed in this study consider all the spectral variables and their inter-correlations, and guide subsequent investigations of the mass spectra and single ion images to allow one to maximize characterization between different regions of any DESI-MS image.

Introduction

Cancer diagnosis is one of the most important challenges in medicine.[1] Important aims in cancer research include enhancing objective assessments and increasing diagnosis accuracy. Wide-ranging research efforts have been focused on understanding the biochemical processes occurring in cancerous cells in relation to those of normal cells in different tissues, particularly as related to changes in lipid, protein and peptide expression.[1–4]

Mass spectrometry (MS) imaging is a powerful analytical technique that allows biomolecules including proteins, peptides and lipids to be analyzed directly from tissue sections with high selectivity. The resulting chemical information is recorded together with spatial distributions on the sample surface.[5] Therefore, MS imaging has the potential to be used to improve the accuracy of histopathological cancer diagnosis by complementing the judgments of expert pathologists[1] with additional objective chemical information.[6] In particular, matrix-assisted laser desorption ionization (MALDI) has been used to determine the distribution of proteins and peptides in biological tissue , as well as some smaller molecules such as phospholipids.[7] More recently, desorption electrospray ionization mass spectrometry (DESI-MS) has been used to investigate lipid profiles in various types of human cancers, such as kidney, bladder, prostate, brain and germ.[4,8–11]These experiments are done in the open laboratory environment and outcomes of such studies suggest that alterations in lipid profiles potentially might serve as biochemical markers of cancer.[12] Remarkably, *in vivo* applications are important future perspectives that are made possible by ambient ionization techniques such as DESI-MS.[13]

As many analytical results are essentially presented as images, processing techniques for representation[14] and interpretation of multi-dimensional data are necessary. An inherent feature of MS hyperspectral image acquisition is the large amount of information contained in each image. This information can be explored and interpreted in a deeper and more efficient way by the application of multivariate techniques.[15–17] Multivariate methods for the treatment of hyperspectral data apply multivariate analysis to the full data matrix, in order to consider both the spatial and the complete spectral information from the samples studied. These methods have been finding application in different fields, including medical diagnostics.[18] It has also been successfully applied to spectroscopic data in order to investigate physiological and pathological changes in living tissue, both animal and human, and to provide information on the health and disease state of tissue.[1,19–23]

Multivariate methods have previously been applied to hyperspectral MS imaging data and showed their potential in different applications.[24–26] In this type of imaging techniques, the hyperspectral data matrix contains, for each pixel of the image, hundreds of variables the values of which constitute the mass spectra.[24]Each spectrum acts as a characteristic fingerprint of the corresponding pixel and can be used to derive information on the chemical composition and properties of the spatial region represented by that pixel. The data matrix – the so-called datacube

– has three dimensions: *X*, *Y* (spatial domains), and *m/z* (spectral domain).[17] Standard multivariate techniques can be applied to the datacube for both exploratory and classification purposes.[18] This approach is advantageous compared to other approaches in which the information contained in the imaging dataset is used by extracting selected individual mass spectra or by averaging the mass spectra of representative areas (which causes the original spatial information to be disregarded) or by plotting the spatial distributions of preselected ions.

In previous studies, a number of unsupervised techniques, such as principal component analysis (PCA) and hierarchical cluster analysis (HCA), were applied to hyperspectral MS imaging data in order to extract information from the mass spectra as related to differentiation of tumor and normal samples.[27] PCA is the technique most commonly used for exploratory investigations of the complex information contained in the full mass spectral dataset and, eventually, for datareduction.[28] HCA allows one to quantify similarities between spectral profiles and to define sample groupings on the basis of similarity values.[29] For example, these techniques have been applied to MALDI-MS imaging data in order to identify areas in the same tissue sample with common spectral patterns and to obtain information about features which are characteristic of each tissue type.[24,29,30] PCA and clustering were also used to confirm characterization of tissues based on disease state, made by pathological evaluation.[30,31] The strategy of selecting some individual spectra of representative pixels in the MALDI image was also used to perform principal component analysis-discriminant analysis (PCA-DA), namely to obtain classification rules based on the PCs as new variables.[32] Application of PCA has also been described for DESI-MS images of tumor and normal tissue sections in different types of human cancers.[4,8] In these cases, PCA was applied to the hyperspectral data using the score values, coded with a color scale, to represent multivariate images of the samples investigated. Such false-color images were compared with the results of pathological evaluation.

Supervised techniques, such as partial least squares-discriminant analysis (PLS-DA), have also been successfully used to discriminate tumor from normal tissue,[33–35] such as kidney and bladder cancers using the information provided by DESI-MS on the tissue lipids profiles.[4,8]

The support vector machine (SVM) approach has also been applied to MS data, using individual mass spectra for each sample to achieve classification.[36] Recently, the application of SVM to MALDI-MS imaging data was proposed for classification of meningiomas in brain tissue.[37] An SVM algorithm has also been used to generate classification models for human brain glioma type, grade and tumor cell concentration based on DESI-MS lipid data.[38]

In this study, an interactive exploratory approach was applied to DESI-MS imaging data of selected pairs of tumor and normal kidney, prostate, germ and bladder tissue sections. The mass spectral data have already been described.[4,8,9,11] A PCA-brushing strategy was used to extract information on the most highly characterizing compounds for tissue sections, while maintaining both the spatial and chemical information of the imaging data. An evaluation of the selectivity of

this multivariate strategy in recognizing the tumor tissue in different types of human cancers was made by correlating the lipid information and pathological evaluation for the same samples.

Interactive multivariate exploration of the spatial and chemical information contained in the datacube can be proposed as a general strategy for exploring and interpreting any DESI-MS images and for characterizing different regions in these images. This can be done as a first step, prior to supervised analysis that requires additional levels of information to be applied. In fact, this strategy may be useful for adapting further multivariate approaches, such as supervised pattern recognition methods, to classification or quantification purposes.

Experimental section

Tissue samples on which MS imaging data were available

This study used published mass spectral imaging data taken on tissue samples (listed in Table S1†). A total of nine pairs for papillary renal cell carcinoma and adjacent normal tissue, seven pairs of clear renal cell carcinoma and adjacent normal tissue,[4] eight pairs of prostate cancer and adjacent normal tissue,[9] 11 pairs of seminoma and adjacent normal tissue[11] and 13 pairs of bladder transitional cell carcinoma and adjacent normal tissue,[8] imaged by DESI-MS, were used. Selection of samples for which data were used in this study was made on the basis of pathological diagnosis. In particular, samples with clear recognition of cancer and normal tissue were used. Tissue section preparation and DESI-MS 2D imaging experiments are described in detail in the ESI†. Note that lipid species were tentatively identified based on collision-induced dissociation (CID) tandem MS experiments previously performed and comparison of the generated product ion spectra with literature data.[4,8,9,11,39]

Multivariate method for analysis of hyperspectral data: PCA and brushing procedure

An in-house program was used to convert the raw MS data files into ASCII files (.txt), and they were then imported into Matlab (The MathWorks, Inc., Natick, USA) for subsequent data processing. This involved assembly into 2D ion images using a color scale normalized to the highest value of absolute intensity for each m/z variable. Interpolation by means of a moving-window cubic spline was used in order to mathematically increase the image resolution.

For each DESI-MS image, the information was coded using the corresponding datacube ($X \cdot Y \cdot m/z$) and PCA was performed. To allow this processing, the datacube was first unfolded to a standard two-dimensional data matrix, whose rows correspond to the pixels and columns correspond to the spectral variables, respectively.[18] In addition, the mass range was reduced to m/z 150–950 for germ samples and to m/z 200–950 for all other samples, thereby excluding uninformative regions of the spectra in which only random noise signal was collected, which represent a non-significant contribution to the total ion current (TIC). Therefore, the hypercubes

were structured with 9600 (for germ samples) and 9000 (for all others) *m*/*z* variables in the *m*/*z* domain. Before performing PCA, the number of spectral variables was reduced (1 : 4) by a consecutive-window averaging algorithm. Then, normalization with respect to the TIC was performed in order to correct for instrumental variability. Neither background correction nor smoothing filters was applied.

For each unfolded data matrix, PCA was performed on column-centered data. The principal components (PCs) can be considered as orthogonal (*i.e.*, uncorrelated) directions in the multidimensional data space that efficiently describes large fractions of the information.[40] The projections of the data objects onto the PCs are called scores, while the importance of each original variable in defining a certain PC is given by the loading coefficient. Both scores and loading values can be represented in two-dimensional scatter plots. Interpretation of the relationships between scores and loadings – in terms of relative directions – is commonly performed by visual examinations of the so-called biplots, which are a superimposition of scores and loadings on the same chart. In this study, scores and loadings – being considerably numerous – have been represented in two separated but contiguous scatter plots, in order to obtain higher readability, but the interpretation is exactly the same as for a biplot.[41,42]

An interactive brushing procedure was performed in order to understand the relationships between the PC space and the image space, connecting chemical and spatial information.[43] In particular, the score plot allows a visual inspection of the pixel distribution in PC space. In the score plot, it is possible to visualize groupings that indicate similarities among pixels, on the basis of the information derived from the mass spectra, and which can be associated with the particular characteristics of the samples analyzed. With the brushing procedure, pixels with similar chemical profiles can be manually selected from the score plot in order to identify correspondences between the groups of points in the PC score plot and particular regions of the DESI-MS image. Subsequently, an examination of both the loading plot and the score plot allows chemical characterization of the highlighted region of the image to be achieved, revealing which *m*/*z* peaks are the most important in defining the pixels under consideration. The ESI† includes a video which shows this.

Multivariate data processing was performed by means of in-house Matlab routines.

Results and discussion

DESI-MS images were acquired as described in the original publications and reviewed in the ESI† using methanol/water (1 : 1) as the spray solvent. PCA was applied to the datacube, after unfolding, in order to extract the greatest amount of useful chemical information. Sample UH9812_03 imaged by DESI-MS contains a papillary RCC tissue section (left side, Fig. 1a) and an adjacent normal tissue section (right side, Fig. 1a). Diagnosis was confirmed by evaluation by an

expert pathologist (Dr Liang Cheng, Department of Pathology and Laboratory Medicine, Indiana University School of Medicine, Indianapolis, IN 46202, USA) of H&E stained serial sections. Correspondence between the chemical and the spatial information (2D image) was achieved by applying the brushing approach. This procedure allows the selection of pixels (typically in the form of a cluster) in the score plot and visualization of the corresponding pixels within the image itself. Fig. 1b and c show the score and loading plots of the two lower-order PCs (PC1 and PC2), respectively. Simultaneous examination of the score and loading plots reveals that the distribution of pixels in PC space is in three predominant directions. The relationship between the score and loading plots is evident from the co-directionality of objects and variables in these two plots. In particular, the pixels highlighted in red – with higher positive values of PC1 and PC2 – represent a region of the image clearly characterized by the presence of an ion *m/z* 885.5 with high relative abundance. Conversely, the pixels in black – at low positive and at negative PC1 values – represent a region of the image clearly characterized by the presence of an ion *m/z* 265.2 with high relative abundance in the mass spectra. Finally, the green pixels – with positive values for PC1 and negative values for PC2 – represent a region of the image clearly characterized by ions such as *m/z* 217.2 and *m/z* 271.2 being present with high relative abundance. In fact, a visual inspection of the loading plot reveals that the loading corresponding to these ions lies in the same spatial directions as the green pixels in the score plot, meaning that those pixels are characterized by a relatively higher intensity of such *m/z* variables.

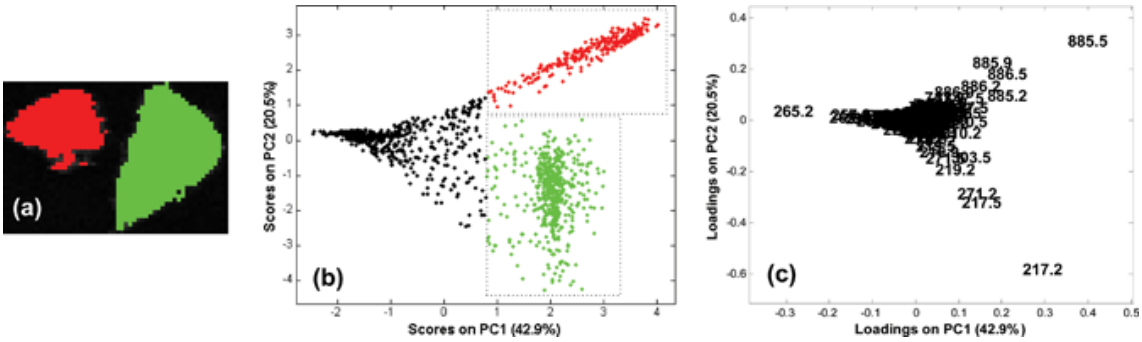

Fig. 1 PCA of renal sample UH9812_03. (a) Correspondence (indicated by matching colors) between PC scores selected by means of the brushing procedure and pixels in the image space (red: tumor section, green: normal section). (b) PC1 (42.9% of the total variance) *vs.* PC2 (20.5% of the total variance) score plot. (c) PC1 *vs.* PC2 loading plot labeled in terms of *m/z* ratio.

The brushing procedure allows one to select a group of points in the score plot and to visualize which pixels they correspond to in image space, where they are highlighted in the same color as in the score plot. As shown in Fig. 1a and b, the red points correlate to the tumor tissue section, while

the green points are related to the normal tissue section. Remarkably, no cross-over was observed using particular PC values selected for renal cell carcinoma, even at the individual pixel level. However, as it will be shown later, this was not true for germ cell carcinoma. Pixels not selected in the score plot (black points) correspond to region of the glass slide not covered with tissue. PC1 (which explains 42.9% of the total data variance) is able to isolate the uncovered glass signals (largely at negative score values) from the signals corresponding to tissue sections; as expected, this difference represents the main information (highest variance) encountered in the total hyperspectral data. Moreover, PC2 (which accounts for 20.5% of the total variance) is able to characterize completely pixels related to the tumor and normal tissue sections, respectively. Analogous pixel distributions and spectral information were found for all the papillary RCC samples examined in this study, with the exception of the ion m/z 271.2 that did not appear consistently in all the samples. In some images, the ion m/z 788.5 also appeared to be important for characterizing background, tumor and normal tissue sections.

The strategy presented supplies a deep interpretation of the information contained in the DESI-MS images. In particular, the examination of the score plot, along with the brushing procedure, allows one to relate pixel structures in the PC space with the spatial domain. The information can be scaled down to the single-pixel level. A simultaneous examination of score plot and loading plot is indispensable to characterize different clusters of pixels on the basis of their chemical information. Examination of score plots alone underutilizes the potential information given by PCA. Moreover, this multivariate strategy considers all the spectral variables and their inter-correlations and can be used to guide subsequent investigations based on individual mass spectra and single ion images, for characterizing different regions of an image. As represented in Fig. 2a–e, the investigation of the spectral domain (m/z) in the original datacube allows individual mass spectra or an average mass spectrum for selected pixels to be visualized. In the $X{\cdot}Y$ spatial domain, ion maps for particular m/z peaks, highlighted by the previous loading analysis, can be created. In particular, Fig. 2c shows the ion map for m/z 885.5 (PI(18 : 0/20 : 4)), while the absolute intensities of other particular lipid species, such as m/z 788.5 (PS(18 : 0/18 : 1)) which is higher in cancerous tissue, m/z 217.2 which is selective for most normal tissue and m/z810.5 (PS(18 : 0/20 : 4)), a non-characterizing species showing approximately equal absolute intensity in cancerous and normal tissues – are shown in Fig. S1 in the ESI †.

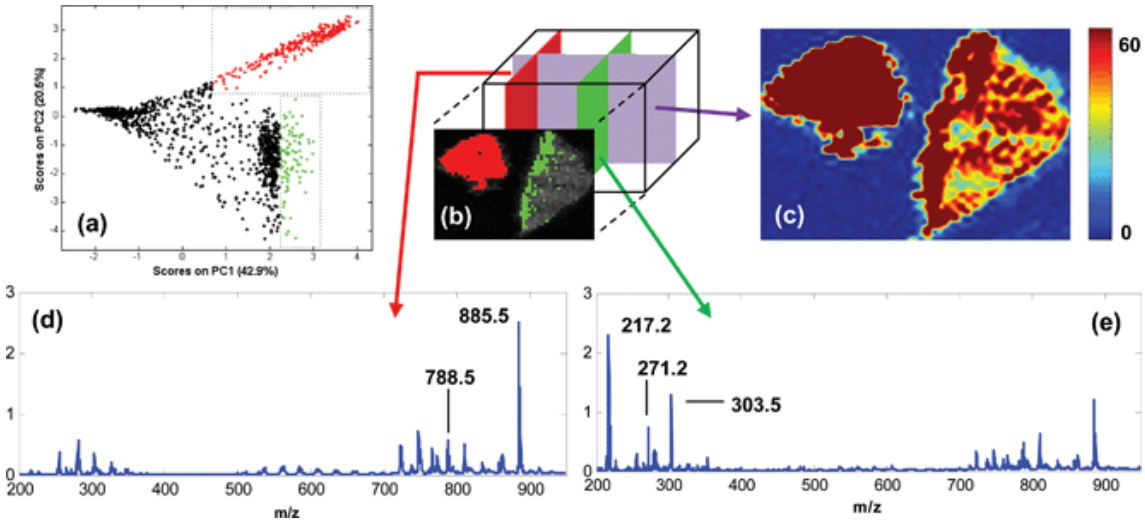

Fig. 2 Interactive hyperspectral analysis for renal sample UH9812_03. (a) PC1 (42.9% of the total variance) *vs.* PC2 (20.5% of the total variance) score plot. (b) Correspondence (indicated by matching colors) between PC scores selected by means of the brushing procedure and pixels in the image space and representation of the data hypercube ($X \cdot Y \cdot m/z$). (c) Ion image of m/z 885.5. (d) Average mass spectrum normalized to the TIC for red pixels. (e) Average mass spectrum normalized to the TIC for green pixels.

The multivariate approach is advantageous compared to that based on absolute intensities in single ion images, since it provides a global visualization of all the chemical information embodied in the data. Furthermore, the strategy followed for normalization of the mass spectra (with respect to the total ion current, TIC) allows unwanted variations due to instrumental variability and/or irregularities in the sample surface – such as folded tissue – to be minimized. For example, the spatial distribution of the green pixels in the PC1 *vs.* PC2 score plot (Fig. 2a), related to the normal tissue section, is different from that of the red pixels and similar to that of other pixels of the normal tissue. A comparison between the average mass spectrum, normalized with the TIC, for red and green pixels (Fig. 2d–e) confirms these outcomes and it is in agreement with the loading plot, previously described. Conversely, the single m/z 885.5 ion map (Fig. 2c) shows similar intensity for red and green pixels, according to the color-code.

The results of these analyses are in agreement with a previous report for this same papillary RCC sample[4] using both visual inspection of the mass spectra and the supervised PLS-DA method, which showed that increased absolute intensities for the lipid species at m/z 885.7, m/z 788.5, m/z 773.5 (PG(18 : 1/18 : 1)) and m/z 913.5 (PI(22 : 4/18 : 0)) occur in cancerous tissue, while the FA species at m/z 215.3 and m/z 217.2 show increased absolute intensity in the normal tissue, correlating inversely with the cancerous tissue.

The same multivariate strategy was applied to DESI-MS images of selected tissue sections of renal clear cell carcinoma, prostate cancer, seminoma and bladder transitional carcinoma. Diagnosis was always confirmed by pathological evaluation of H&E stained serial sections. For the clear cell RCC tissue samples, in contrast to the papillary RCC, no clear visual changes between the absolute intensities of the lipid species for either the cancerous or normal tissue were observable by simple inspection of the mass spectra or by single ion images for many samples, as shown for example in Fig. S2† for sample MH0111_10. This emphasizes the need for a multivariate exploratory analysis that uses the entirety of the mass spectral data for characterizing different regions of the DESI-MS images. Applying PCA and the brushing procedure to the hyperspectral image, the score plot and the image with pixels colored on the basis of brushing selections (Fig. 3a and b) show that, along the first principal component, it is possible to isolate the information on the background (at positive score values) from that related to the tissue sections. Conversely, PC2 is able to completely separate pixels related to tumor (left section, Fig. 3a) and normal (right section, Fig. 3a) tissues, indicated as red and green pixels respectively. The most important ions in the definition of PC2 appear to be *m*/*z* 281.5, *m*/*z* 885.5, *m*/*z* 303.2 and *m*/*z* 562.9, as it can be observed in the PC1 *vs.* PC2 loading plot (Fig. 3c). It can be noted that, for this type of cancer, it was not possible to achieve a complete pixel differentiation between background, tumor and normal tissue sections for all the samples investigated when applying the same multivariate strategy (see Table 1 in the ESI†).

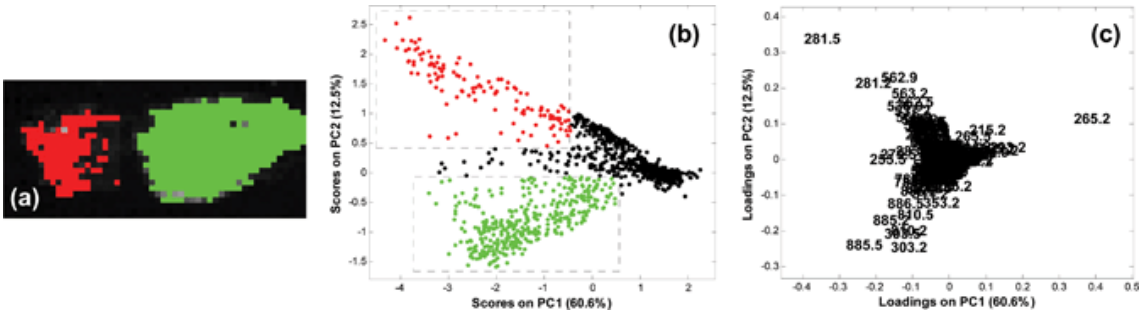

Fig. 3 PCA for renal sample MH0111_10. (a) Correspondence (indicated by matching colors) between PC scores selected by means of the brushing procedure and pixels in the image space (red: tumor section, green: normal section). (b) PC1 (60.6% of the total variance) *vs.* PC2 (12.5% of the total variance) score plot. (c) PC1 *vs.* PC2 loading plot labeled in terms of *m*/*z* ratios.

In the case of prostate cancer, visual inspection of the mass spectra and univariate ion images reveals that cholesterol sulfate (CS, *m*/*z* 465.5) is present in high relative abundance in precancerous lesions or cancerous tissue, while it is undetected in adjacent normal tissue.[9] The sample UH0002_25 is an example of cancerous prostate tissue (left section, Fig. 5e) showing high

relative intensities of CS in regions with high concentration of tumor cells, while the CS signal in normal tissue (right section, Fig. 5e) is almost absent. From PCA, the ion *m*/*z* 465.5 appears to be important in PC2 definition, together with the ions *m*/*z* 885.5 and *m*/*z* 788.5, as shown in Fig. 4. Along this axis, it is possible to visualize the separation between pixels related to tumor and normal tissue. By contrast, the first PC explains the difference between pixels of tissue sections and those related to the uncovered glass slide. Finally, it is important to highlight that pixels of tumor section can be almost completely separated from those of normal tissue in PC1 *vs.* PC2 score space. Less efficient recognition can be achieved in the single ion images which consider the ions *m*/*z* 465.5, *m*/*z* 788.5 and *m*/*z* 885.5 individually (Fig. 5b–d). This confirms once again that the multivariate strategy permits to extract the highest amount of useful information, offering a global chemical characterization of the samples under study.

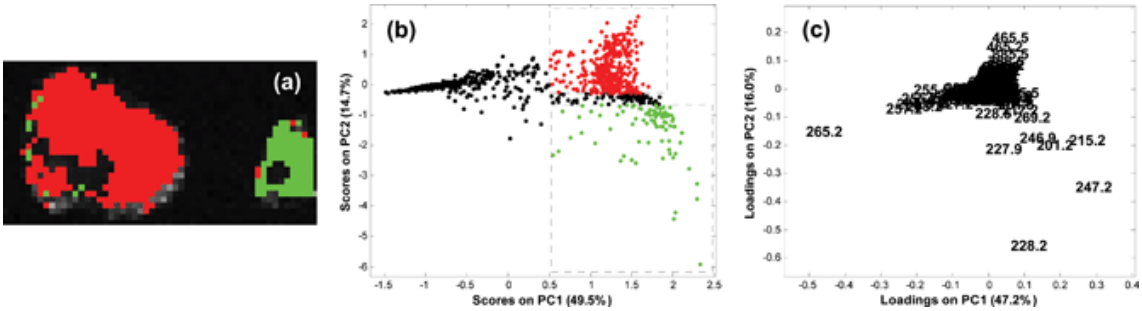

Fig. 4 PCA for prostate sample UH0002_25. (a) Correspondence (indicated by matching colors) between PC scores selected by means of the brushing procedure and pixels in the image space (red: tumor section, green: normal section). (b) PC1 (47.2% of the total variance) *vs.* PC2 (16.0% of the total variance) score plot. (c) PC1 *vs.* PC2 loading plot labeled in terms of *m*/*z* ratios.

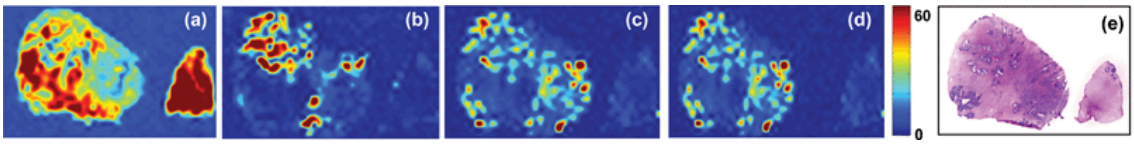

Fig. 5 Negative ion mode imaging of prostate tumor tissue and adjacent normal tissue of sample UH0002_25. (a) Ion image of *m*/*z* 215.2. (b) Ion image of *m*/*z*465.5. (c) Ion image of *m*/*z* 788.5. (d) Ion image of *m*/*z* 885.5. (e) H&E stained tissue sections of the tumor tissue (left) and normal tissue (right).

By examining Fig. 4a, one notices that the manual pixel selection applied, resulted in some pixel misrecognitions on the perimeter of the tissue sections. In particular, some pixels on the border of the normal tissue are erroneously colored in red. This would have been expected for

tumor tissue, but the pathological evaluation indicates no presence of tumor in this sample section. A few similar misrecognitions occurred in some other samples, often on the borders of normal tissue sections. This may suggest that contaminations on the borders might have occurred during tissue cutting/handling. Nevertheless, misleading chemical differences are not unexpected between the major surfaces of tissue section and the marginal region due to differences in the exposure history of the original sectioned tissue and that of the tissue sections subsequently created during cutting/handling operations.

For germ carcinoma, differently from any other tissue analyzed, mass spectra and single ion images show that seminolipid species are found exclusively in normal tubules tissue, while they are absent in normal fat or muscles tissue. For instance, sample UH0201_09 consists of a normal tubules tissue section (right, Fig. S3a†), which presents a high abundance of m/z 795.5 (seminolipid 16 : 0/16 : 0), and a seminoma tissue section (left, Fig. S3b†) in which this species is completely absent, while higher intensities of the ions m/z 885.5, m/z 788.5 and m/z 175.2 (ascorbic acid) are mainly observed.[11] In Fig. S3†, the score plot and the image with pixels colored on the basis of brushing selections show that it is possible to clearly separate information related to the normal tissue (green points), tumor tissue (red points), and background (non-selected black points) in the PC1 *vs.* PC2 score space. The loading analysis (Fig. S3c†) confirms that the seminolipid m/z 795.5 has a higher relative intensity in the pixels that correspond to normal tissue. Conversely, pixels corresponding to tumor tissue are characterized by a higher relative abundance of the ions m/z 281.5, m/z 175.2, m/z 788.5 and m/z 885.5. The last three ions are more important for achieving characterization of tumor tissue and normal fat or muscles tissue, in which the absence of seminolipids is observed, as shown in Fig. S4† for sample UH0201_20.

Bladder transitional cell carcinoma tissue (left section) and adjacent normal tissue (right section) are compared in Fig. S5a† (sample UH0208_18). Also in this case, the PC1 *vs.* PC2 score space (Fig. S5b†) allows characterization of background (black points), tumor tissue (red points), and normal tissue (green points). From loading analysis (Fig. S5c†), the most important ions for the characterization of tumor tissue appear to be m/z 563.2 and m/z 281.9, while theglycerophospholipid (GP) species PS(36 : 1) and PI(38 : 4) are found to be relevant for characterizing normal tissue. Increased intensities of free FAs and their dimers were observed consistently as the main contributors to the characterization of normal and cancerous tissue for the majority of the samples analyzed, while no relevant information can be observed by visual inspection of the m/z 788.5 and m/z 885.5 ion images. These outcomes are in agreement with previous observations that fatty acids and dimers are the main species responsible for characterization of cancer and normal bladder tissue by DESI-MS imaging.[8]

Finally, according to the information in the loading plots, it can be noted that in different types of human tissue sections, some common ions, such as m/z885.5, m/z 788.5, free FAs and their dimers, appear to be important for characterizing between cancer and normal tissue. Other ions,

such as *m/z* 217.2 for kidney, *m/z* 465.5 for prostate tissue and *m/z* 795.5 for germ tissue, are selective for particular types of tissue, confirming that the information of the entire lipid profile can reflect the selectivity of the tissue type and not only the differences between neoplastic and normal tissue.

Conclusions and perspectives

MS imaging is a powerful tool that allows lipid species to be detected directly from tissue sections, thus giving information on both lipid characterization and spatial distribution on sample surfaces. This chemical and spatial information should be processed by means of multivariate analysis.

In this study, an interactive PCA-brushing approach was applied for the first time to hyperspectral DESI-MS imaging data in order to exploit the chemical information provided by mass spectrometry, allowing tissue characterization that may be valuable in cancer diagnosis and complementary to the traditional histopathological examination. Exploratory analysis is the first step necessary for visualizing and understanding the information contained in datacubes. In more detail, it allows to perform a thorough chemical characterization of a sample, understanding relations between any portion of the sample image and its chemical composition – thanks to the combined pixel-score-loading analysis. Data taken for more than 40 samples of tumor and normal kidney, prostate, germ and bladder tissue were investigated. Generally, the exploration of the PC1 *vs.* PC2 score space showed clusters of pixels related to uncovered glass slide, tumor and normal tissue sections, respectively. At the same time, the PC1 *vs.* PC2 loading plot showed consistently that the *m/z* variables important for characterizing tumor and normal tissues of a particular organ were the same in all the different samples considered. These outcomes confirm that an interactive strategy presented can be a tool of general usefulness for the multivariate chemical characterization of biological tissues on the basis of hyperspectral MS imaging data. The strategy presented can be also used to guide subsequent investigations of the mass spectra and single ion images for characterizing different regions of any DESI-MS images which can be interpreted according to additional level of information, such as the assignment of health and disease state of tissue sections as described in this study. Importantly, the PC scores can be used as new variables that may allow efficient applications of more complex chemometric strategies for supervised classification.

Finally, we envision that applying such interactive hyperspectral approaches on DESI-MS imaging data collected using a “morphologically friendly” (nondestructive) solvent system[6] will allow stronger correlations between the chemical information – related pixel by pixel to the spatial information by applying brushing procedure – and morphological information. In consequence, deeper interpretation of the outcomes should be achieved.

Acknowledgements

This work was supported by the Indiana Clinical and Translational Sciences Institute (in part by Grant # RR025761 from the NIH, National Center for Research Resources, Clinical and Translational Sciences Award). We gratefully acknowledge Dr Timothy Masterson at IU School of Medicine for assistance in obtaining the original tissue samples, Dr Liang Cheng for performing pathological evaluations, the Purdue University Center for Cancer Research and its director Dr Timothy Ratliff for assistance and support, and Marco Vincenti, professor of Analytical Chemistry at the University of Turin, Italy.

References

1. H. Akbari, K. Uto, Y. Kosugi, K. Kojima and N. Tanaka, *Cancer Sci.*, 2011, **102**, 852–857 CrossRef CAS Search PubMed .
2. H. Bateson, S. Saleem, P. M. Loadman and C. W. Sutton, *J. Pharmacol. Toxicol. Methods*, 2011, **64**, 197–206 CrossRef CAS Search PubMed .
3. D. T. Dicker, J. Lerner, P. Van Belle, S. F. Barth, D. Guerry, M. Herlyn and W. S. El-Deiry, *Cancer Biol. Ther.*, 2006, **5**, 1033–1038 CrossRef Search PubMed .
4. A. L. Dill, L. S. Eberlin, C. Zheng, A. B. Costa, D. R. Ifa, L. A. Cheng, T. A. Masterson, M. O. Koch, O. Vitek and R. G. Cooks, *Anal. Bioanal. Chem.*, 2010, **398**, 2969–2978 CrossRef CAS Search PubMed .
5. E. R. Amstalden van Hove, D. F. Smith and R. M. A. Heeren, *J. Chromatogr., A*, 2010, **1217**, 3946–3954 CrossRef CAS Search PubMed .
6. L. S. Eberlin, C. R. Ferreira, A. L. Dill, D. R. Ifa, L. Cheng and R. G. Cooks, *ChemBioChem*, 2011, **12**, 2129–2132 CrossRef CAS Search PubMed .
7. L. S. Eberlin, C. R. Ferreira, A. L. Dill, D. R. Ifa and R. G. Cooks, *Biochim. Biophys. Acta*, 2011, **1811**, 946–960 CAS Search PubMed .
8. A. L. Dill, L. S. Eberlin, A. B. Costa, C. Zheng, D. R. Ifa, L. A. Cheng, T. A. Masterson, M. O. Koch, O. Vitek and R. G. Cooks, *Chem.–Eur. J.*, 2011, **17**, 2897–2902 CrossRef CAS Search PubMed .
9. L. S. Eberlin, A. L. Dill, A. B. Costa, D. R. Ifa, L. Cheng, T. Masterson, M. Koch, T. L. Ratliff and R. G. Cooks, *Anal. Chem.*, 2010, **82**, 3430–3434 CrossRef CAS Search PubMed .
10. L. S. Eberlin, A. L. Dill, A. J. Golby, K. L. Ligon, J. M. Wiseman, R. G. Cooks and N. Y. R. Agar, *Angew. Chem., Int. Ed.*, 2010, **49**, 5953–5956 CrossRef CAS Search PubMed .
11. T. A. Masterson, A. L. Dill, L. S. Eberlin, M. Mattarozzi, L. Cheng, S. D. W. Beck, F. Bianchi and R. G. Cooks, *J. Am. Soc. Mass Spectrom.*, 2011,**22**, 1326–1333 CrossRef CAS Search PubMed .
12. P. Waloszczyk, T. Janus, J. Alchimowicz, T. Grodzki and K. Borowiak, *Diagn. Pathol.*, 2011, **6**, 22 CrossRef Search PubMed .

13. P. Nemes and A. Vertes, *TrAC, Trends Anal. Chem.*, 2012, **34**, 22–34 CrossRef CAS Search PubMed .
14. P. K. Dasgupta, Y. J. Chen, C. A. Serrano, G. Guiochon, H. H. Liu, J. N. Fairchild and R. A. Shalliker, *Anal. Chem.*, 2010, **82**, 10143–10150 CrossRef CAS Search PubMed .
15. S. J. Blanksby and T. W. Mitchell, in *Annual Review of Analytical Chemistry*, ed. E. S. Z. R. N. Yeung, 2010, vol. 3, pp. 433–465 Search PubMed .
16. R. Bro, J. J. Workman, P. R. Mobley and B. R. Kowalski, *Appl. Spectrosc. Rev.*, 1997, **32**, 237–261 CrossRef CAS Search PubMed .
17. L. A. Klerk, A. Broersen, I. W. Fletcher, R. van Liere and R. M. A. Heeren, *Int. J. Mass Spectrom.*, 2007, **260**, 222–236 CrossRef CAS Search PubMed .
18. J. Burger and A. Gowen, *Chemom. Intell. Lab. Syst.*, 2011, **108**, 13–22 CrossRef CAS Search PubMed .
19. D. T. Dicker, J. M. Lerner and W. S. El-Deiry, *Cell Cycle*, 2007, **6**, 2563–2570 CrossRef CAS Search PubMed .
20. S. Duraipandian, W. Zheng, J. Ng, J. J. H. Low, A. Ilancheran and Z. Huang, *Analyst*, 2011, **136**, 4328–4336 RSC .
21. M. E. Martin, M. B. Wabuyele, K. Chen, P. Kasili, M. Panjehpour, M. Phan, B. Overholt, G. Cunningham, D. Wilson, R. C. DeNovo and T. Vo-Dinh,*Ann. Biomed. Eng.*, 2006, **34**, 1061–1068 CrossRef Search PubMed .
22. S. V. Panasyuk, S. Yang, D. V. Faller, D. Ngo, R. A. Lew, J. E. Freeman and A. E. Rogers, *Cancer Biol. Ther.*, 2007, **6**, 439–446 CrossRef Search PubMed .
23. A. M. Siddiqi, H. Li, F. Faruque, W. Williams, K. Lai, M. Hughson, S. Bigler, J. Beach and W. Johnson, *Cancer Cytopathol.*, 2008, **114**, 13–21 Search PubMed .
24. T. Alexandrov and J. H. Kobarg, *Bioinformatics*, 2011, **27**, I230–I238 CrossRef CAS Search PubMed .
25. B. J. Tyler, G. Rayal and D. G. Castner, *Biomaterials*, 2007, **28**(15), 2412–2423 CrossRef CAS Search PubMed .
26. V. S. Smentkowski, S. G. Ostrowski, F. Kollmer, A. Schnieders, M. R. Keenan, J. A. Ohlhausen and P. G. Kotula, *Surf. Interface Anal.*, 2008, **40**(8), 1176–1182 CrossRef CAS Search PubMed .
27. L. Wu, J. S. Felton and K. J. Wu, *Methods Mol. Biol.*, 2010, **656**, 267–281 CAS Search PubMed .
28. J. A. Fernandez, B. Ochoa, O. Fresnedo, M. T. Giralt and R. Rodriguez-Puertas, *Anal. Bioanal. Chem.*, 2011, **401**, 29–51 CrossRef CAS Search PubMed .
29. D. Bonnel, R. Longuespee, J. Franck, M. Roudbaraki, P. Gosset, R. Day, M. Salzet and I. Fournier, *Anal. Bioanal. Chem.*, 2011, **401**, 149–165 CrossRef CAS Search PubMed .

30. S. O. Deininger, M. P. Ebert, A. Fuetterer, M. Gerhard and C. Roecken, *J. Proteome Res.*, 2008, **7**, 5230–5236 CrossRef CAS Search PubMed .
31. S. Kang, H. S. Shim, J. S. Lee, D. S. Kim, H. Y. Kim, S. H. Hong, P. S. Kim, J. H. Yoon and N. H. Cho, *J. Proteome Res.*, 2010, **9**, 1157–1164 CrossRef CAS Search PubMed .
32. M.-C. Djidja, E. Claude, M. F. Snel, S. Francese, P. Scriven, V. Carolan and M. R. Clench, *Anal. Bioanal. Chem.*, 2010, **397**, 587–601 CrossRef CAS Search PubMed .
33. H. Gu, Z. Pan, B. Xi, V. Asiago, B. Musselman and D. Raftery, *Anal. Chim. Acta*, 2011, **686**, 57–63 CrossRef CAS Search PubMed .
34. K.-C. Schaefer, J. Balog, T. Szaniszlo, D. Szalay, G. Mezey, J. Denes, L. Bognar, M. Oertel and Z. Takats, *Anal. Chem.*, 2011, **83**, 7729–7735 CrossRef Search PubMed .
35. O. P. Whelehan, M. E. Earll, E. Johansson, M. Toft and L. Eriksson, *Chemom. Intell. Lab. Syst.*, 2006, **84**, 82–87 CrossRef CAS Search PubMed .
36. M. Ceccarelli, A. d'Acierno and A. Facchiano, *BMC Bioinf.*, 2009, **10**, 59 CrossRef Search PubMed .
37. N. Y. R. Agar, J. G. Malcolm, V. Mohan, H. W. Yang, M. D. Johnson, A. Tannenbaum, J. N. Agar and P. M. Blacks, *Anal. Chem.*, 2010, **82**, 2621–2625 CrossRef CAS Search PubMed .
38. L. S. Eberlin, I. Norton, A. L. Dill, A. J. Golby, K. L. Ligon, S. Santagata, R. G. Cooks and N. Y. R. Agar, *Cancer Res.*, 2012, **72**(3), 645–654 CrossRef CAS Search PubMed .
39. M. J. Cole and C. G. Enke, *Anal. Chem.*, 1991, **63**, 1032–1038 CrossRef CAS Search PubMed .
40. T. I. Jolliffe, *Principal Component Analysis*, Springer, New York, 2002 Search PubMed .
41. J. C. Gower and D. J. Hand, *Biplots*, Chapman & Hall, London, 1996 Search PubMed .
42. U. Kohler and M. Luniak, *Stata J.*, 2005, **5**(2), 208–223 Search PubMed .
43. H. Grahn and P. Geladi, *Techniques and Applications of Hyperspectral Image Analysis*, Wiley, New York, 2007 Search PubMed .